DTI Based Diagnostic Prediction of a Disease via Pattern Classification

Madhura Ingalhalikar¹, Stathis Kanterakis¹, Ruben Gur²,
Timothy P.L. Roberts^{3,*}, and Ragini Verma¹

¹ Section of Biomedical Image Analysis,
University of Pennsylvania, Philadelphia, PA, USA

{Madhura.Ingalhalikar,Stathis.Kanterakis,Ragini.Verma}@uphs.upenn.edu

² Brain Behavior Laboratory, University of Pennsylvania, Philadelphia, PA, USA
gur@bbl.med.upenn.edu

³ Center for Autism Research, Children's Hospital of Philadelphia,
Philadelphia, PA, USA
robertstim@email.chop.edu

Abstract. The paper presents a method of creating abnormality classifiers learned from Diffusion Tensor Imaging (DTI) data of a population of patients and controls. The score produced by the classifier can be used to aid in diagnosis as it quantifies the degree of pathology. Using anatomically meaningful features computed from the DTI data we train a non-linear support vector machine (SVM) pattern classifier. The method begins with high dimensional elastic registration of DT images followed by a feature extraction step that involves creating a feature by concatenating average anisotropy and diffusivity values in anatomically meaningful regions. Feature selection is performed via a mutual information based technique followed by sequential elimination of the features. A non-linear SVM classifier is then constructed by training on the selected features. The classifier assigns each test subject with a probabilistic abnormality score that indicates the extent of pathology. In this study, abnormality classifiers were created for two populations; one consisting of schizophrenia patients (SCZ) and the other with individuals with autism spectrum disorder (ASD). A clear distinction between the SCZ patients and controls was achieved with 90.62% accuracy while for individuals with ASD, 89.58% classification accuracy was obtained. The abnormality scores clearly separate the groups and the high classification accuracy indicates the prospect of using the scores as a diagnostic and prognostic marker.

1 Introduction

High dimensional pattern classification methods like support vector machine (SVM) can be used to capture multivariate relationships among various anatomical regions for more effectively characterizing group differences as well as quantifying the degree of pathological abnormality associated with each individual.

* The authors would like to acknowledge support from the NIH grants: MH079938, DC008871 and MH060722.

In this paper, we propose a method for region-based pattern classification that creates abnormality classifiers using information from DTI data in anatomically meaningful regions.

Classification not only elucidates regions that are sufficiently affected by pathology so as to be able to differentiate the group from that of healthy controls, but also assigns each individual with a probabilistic score indicating the degree of abnormality. Such a score can be used as a diagnostic tool to predict the extent of disease or group difference and assess progression or treatment effects.

Pattern classification has been previously implemented in neuroimaging studies, mainly in structural imaging [1,2,3,4,5] and a few using DTI information [6,7]. The process can be divided into 4 steps: feature extraction, feature selection and classifier training and testing or cross-validation. Pattern classification methods differ on the basis of *features extracted*. Typical methods for feature extraction use direct features like structural volumes and shapes [2,5], principal component analysis (PCA) [6], automated segmentation [4], wavelet decomposition [1], and Bayes error estimation [7]. High dimensionality of the feature space and limited number of samples pose a significant challenge in classification [2]. To solve this problem, *feature selection* is performed in order to produce a small number of effective features for efficient classification and to increase the generalizability of the classifier. Feature selection methods that are commonly used for structural neuro-imaging data involve filtering methods (e.g. ranking based on Pearson's correlation coefficient) and/or wrapper techniques (e.g. recursive feature elimination (RFE)). Classifiers are then trained on the selected features either using linear classifiers like linear discriminant analysis (LDA) or non-linear classifiers like k-nearest neighbors (kNN)[7] and support vector machines (SVM) [1,2,4,5]. Prior to their application to a new test subject, it is important that the classifier be cross-validated, for which jack-knifing (leave-one-out) is a popular technique. When applied to any subject individual, these cross-validated classifiers produce an abnormality score that can be combined with clinical scores to aid in diagnosis and prognosis.

While most classification work used structural images few studies did attempt DTI-based classification. Caan et al. [6] used fractional anisotropy (FA) and linear anisotropy images as their features, followed by dimensionality reduction using PCA and trained the data using linear discriminant analysis (LDA). Wang et al. [7] used a k-NN classifier trained on full brain volume FA and geometry maps. The former used a linear classifier incapable of capturing complex non-linear relationships, while the latter used a non-linear classification albeit on the full brain.

We describe a classification methodology based on DTI features of anisotropy and diffusivity computed from anatomically meaningful regions of interest (ROI). The regions are based on an atlas and do not require a priori knowledge of affected region. An efficient two step feature ranking and selection technique is implemented. A non-linear SVM classifier is then trained and cross-validated using the leave-one-out paradigm. The classifiers are then applied to 2 different

datasets: one includes patients with SCZ and another with ASD, where each individual is assigned a classification score that quantifies degree of pathology relative to the population. Despite the difference in size of the dataset and the heterogeneity of the pathology, we obtained classifiers with high cross-validation accuracy. This establishes the applicability of the classifier scores for aiding diagnosis and prognosis. The classifier also produces a ranking of features, anatomical regions in our case, which contribute towards the separation of groups, thereby making the classification score of abnormality clinically meaningful.

2 Methods

Our method of creating ROI-based classifiers using DTI-based information involves 4 steps: a. Feature extraction b. Feature selection c. Classifier training and d. Cross-validation.

2.1 Feature Extraction

Our method begins with spatial normalization of all the tensor images to a standard atlas created by Wakana et al. [8], with 176 anatomical regions labeled. A high dimensional elastic registration introduced that involves matching tensor attributes, is used and all scans are registered to the template [9]. In the earlier studies that involved classification [6,7], the whole brain was used as a feature. This is challenging for training classifiers as the sample size is small while the data dimensionality is very high. As a result of the inter-individual structural variability, using fewer features computed from small anatomical regions of interest may lower the dimensionality while maintaining the spatial context, thereby producing more robust features [4]. Therefore we use the ROI's from Wakana et al.'s [8], diffusion tensor template, as shown in figure 1(a). The FA and mean diffusivity (TR) are computed from the spatially normalized tensor images for each subject and then averaged over each ROI. Thus, each subject is associated with a feature vector (eq. 1) that involves 'n' ROI's. Eq. 1 describes the feature vector where each component is the average value of the scalar in that ROI.

$$f_s = (ROI_1^{FA}, ROI_2^{FA} \dots ROI_n^{FA}, ROI_1^{Tr}, ROI_2^{Tr} \dots ROI_n^{Tr}) \quad (1)$$

2.2 Feature Selection

Identifying the most characteristic features is critical for minimizing the classification error. This can be achieved by ranking and selection of the relevant features and eliminating all the redundant features. We use a two step feature ranking and selection as suggested in most of the pattern classification theories, where the two steps involve a filter and a wrapper yielding best results [10]. We filter the features using minimum redundancy and maximum relevance (mRMR) ranking introduced by Peng et al. [11]. This method involves computation of the mutual information between 2 variables x and y which is defined by their probability density functions $p(x)$, $p(y)$ and $p(x, y)$.

$$I(x; y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (2)$$

Based on the mutual information values, Peng et al. have defined 2 criteria to find the near optimal set of features: 1. Maximum relevance and 2. Minimum redundancy. In maximum relevance, the selected feature x_i is required individually to have the largest mutual information with the class label c which reflects the highest dependency on the target class. If we assume the optimal subset of features is S that consists of m features, the maximum relevance of the whole dataset is obtained by $\max(D(S, c))$ where D is given by equation 3.

$$D = \frac{1}{|S|^2} \sum I(x_i; c) \quad (3)$$

It is likely that features with maximum relevance are highly redundant i.e. the features highly depend upon each other. The minimum redundancy criteria therefore is given by $\min(R(S))$ where R is obtained from equation 4.

$$R = \frac{1}{|S|^2} \sum I(x_i; x_j) \quad (4)$$

Optimizing D and R values simultaneously, we get the ideal set of features that can be implemented in classification of the data. Details of the mRMR method can be found in [11]. For selecting the feature subset after ranking, the cross validation error from the classifier is computed sequentially for the ranked features and plotted against the number of features. From this error plot, the area with consistently lowest error is marked and the smallest number of features from that area are chosen as the candidate feature set [11].

Once a feature subset is chosen from the mutual information based filter, a recursive backward elimination (RFE) wrapper is applied over this subset. Unlike the filter, the wrapper explicitly minimizes the classification error by recursively excluding one feature with the constraint that the resultant feature set has lower leave one out (LOO) error [12].

2.3 Training Classifiers Using SVM

A non-linear support vector machine (SVM) is amongst the most powerful pattern classification algorithms, as it can obtain maximal generalization when predicting the classification of previously unseen data compared to other nonlinear classifiers [13]. By using a kernel function, it maps the original features into higher dimensional space where it computes a hyperplane that maximizes the distance from the hyperplane to the examples in each class. Having found such a hyperplane, the SVM can then predict the classification of an unlabeled example by mapping it into the feature space and checking on which side of the separating plane the example lies.

The input to the classifier consists of feature matrix and a vector of class labels consisting of 1 and -1 values defining the two classes (the controls are labeled

as 1 and the patients as -1, in our case). The classifier performs a non-linear mapping $\phi : R^d \rightarrow H$ by using a kernel function, which maps the feature space R^d to a higher dimensional space H . We use the Gaussian radial basis function as a kernel function.

Based on the distance from the hyperplane, the classifier computes a probabilistic score between 1 and -1 for each test subject. When the probabilistic score is ≥ 0 , the subject is classified as class 1 (controls), otherwise as class 2 (patients). The probabilistic classifier score, therefore, represents the level of abnormality in the subject.

2.4 Cross-Validation

We compute the probabilistic abnormality score for each subject by implementing the leave one out (LOO) cross validation method. In the LOO validation, one sample is chosen for testing, while other samples are trained using the methods described in section 2.3. The classifier is evaluated based on the classification result of the test subject. By repeatedly leaving each subject out as a test subject, obtaining its abnormality score, and averaging over all the left-out subjects we obtain the average classification rate. Validation is also performed by plotting the receiver operating characteristic (ROC) curves. The ROC is a plot of sensitivity vs. (1-specificity) of the classifier when the discrimination threshold is varied.

3 Results

Two datasets were used for carrying out the classification:

- The SCZ dataset consisted of 37 female controls and 27 female SCZ patients. The DWI images were acquired on Siemens 3T scanner with $b=1000$ and 64 gradient directions.
- The ASD dataset consisted of 25 children with ASD and 23 typically developing kids. The DWI images were acquired on Siemens 3T scanner with $b=1000$ and 30 gradient directions.

Fig. 1. (a) ROI's introduced by Wakana et al. [8] showing 176 structures

Fig. 2. Plot showing LOO classification (abnormality) scores vs. frequency of occurrence of the score value (PDF) where green indicates patients and blue indicates controls. Each subject’s score is represented with a star. (a) SCZ data where the stars overlap as a result of a large number of samples. The controls and patients are very well separated with 90.62% accuracy. (b) ASD dataset shows 89.58% classification even with less number of samples.

The SCZ dataset has a large sample size in comparison with the ASD dataset. Moreover, the SCZ data consists of a cohesive population as only females are considered, while the ASD data consists of a heterogeneous population. We chose such different datasets to test the versatility of our classification method.

As mentioned in section 2, the scans were spatially normalized to the DTI template. Subsequently, the mean FA and MD measures for each ROI were derived. Feature selection was performed using mutual information values and the RFE wrapper as described in section 2.2. After employing the filter, 130 top features for SCZ dataset were chosen while for ASD 150 top features were picked. The selection of number of features was based on the error plot as described in section 2.2. From this feature subset, the RFE wrapper chose 25 top features for SCZ and 17 top features for ASD by minimizing the classification error. For the non-linear classifier, a suitable kernel size was determined based on the ROC curves, after testing different σ values ranging from 0.001 to 1. It was found that the kernel size ranging from 0.05 - 0.15 was optimal. The C value, which is the trade off parameter between training error and SVM margin, was tested for 1,5,7,10 and 50, and it was observed that values between 5 and 50 were a better choice for our classification.

Using the LOO method, we computed the classifier score for each subject. Figure 2 shows the classification results. A normal probability density (PDF) which represents the likelihood of each LOO score is plotted against the LOO score. The average LOO classification for SCZ data was 90.62% (6 subjects misclassified) while for ASD data it was 89.58% (5 subjects misclassified). The ROC curves are shown in figure 3. Both the datasets shows a steep curve with a large area under the curve (AUC) of 0.905 for SCZ and 0.913 for ASD (fig. 3(a and b)). We also applied our method to test asymptomatic relatives of SCZ patients. The abnormality scores of the relatives were closer to the patients than the healthy controls, suggesting similar brain abnormality patterns.

Fig. 3. ROC curves for classification between controls and patients. The area under the curve (AUC) defines the level of performance. (a) SCZ data with an AUC of 0.905 (b) ASD with an AUC of 0.913.

Fig. 4. Top ranked ROI's mapped on template image. These ROI's contributed largely towards classification based two step feature selection.(a) SCZ (b) ASD data.

Besides computing the abnormality score, it is important to know which areas of the brain significantly contribute towards the classification. The selected features are the most discriminating features and contribute most towards group differences between controls and patients. Fig. 4 shows the top 25 ROI's for SCZ (fig. 4(a)) and 17 ROI's for ASD (fig. 4(b)) that were chosen by the wrapper are overlaid on the template image and color coded according to their rank.

4 Conclusion

We have presented a classification methodology based on anatomically meaningful DTI features for identification of white matter abnormality in diseases like SCZ and ASD. Use of ROI based features reduces dimensionality and simplifies the clinical interpretation of pathology induced changes. Superior experimental results over different populations indicate that our method can successfully classify patients and controls, despite the sample size and heterogeneity of the population. Based on the abnormality score of the test subject, this method can quantify the degree of pathology and the score can be potentially used as a clinical biomarker.

Thus, our method can be applied in various studies involving disease progression, treatment effects etc. The application to family members underlines the applicability of the method to prognosis. Future work includes extending the classification method for controlling for age and treatment effects. Moreover, we can add other features like radial and axial diffusivity measures or utilize the whole tensor in log-euclidean form as our feature for further investigation.

References

1. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage* 21(1), 46–57 (2004)
2. Golland, P., Grimson, W.E.L., Shenton, M.E., Kikinis, R.: Detection and analysis of statistical differences in anatomical shape. *Med. Image Anal.* 9(1), 69–86 (2005)
3. Yushkevich, P., Joshi, S., Pizer, S.M., Csernansky, J.G., Wang, L.E.: Feature selection for shape-based classification of biological objects. *Inf. Process Med. Imaging* 18, 114–125 (2003)
4. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: Compare: classification of morphological patterns using adaptive regional elements. *IEEE Trans Med. Imaging* 26(1), 93–105 (2007)
5. Pohl, K.M., Sabuncu, M.R.: A unified framework for mr based disease classification. *Inf. Process Med. Imaging* 21, 300–313 (2009)
6. Caan, M.W.A., Vermeer, K.A., van Vliet, L.J., Majoie, C.B.L.M., Peters, B.D., den Heeten, G.J., Vos, F.M.: Shaving diffusion tensor images in discriminant analysis: a study into schizophrenia. *Med. Image Anal.* 10(6), 841–849 (2006)
7. Wang, P., Verma, R.: On classifying disease-induced patterns in the brain using diffusion tensor images. *Med. Image Comput. Comput. Assist Interv.* 11(Pt 1), 908–916 (2008)
8. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C.M., Mori, S.: Fiber Tract-based Atlas of Human White Matter Anatomy. *Radiology* 230(1), 77–87 (2004)
9. Ingallhalikar, M., Yang, J., Davatzikos, C., Verma, R.: Dti-droid: Diffusion tensor imaging-deformable registration using orientation and intensity descriptors. *International Journal of Imaging Systems and Technology* 20(2), 99–107 (2010)
10. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182 (2003)
11. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal. Mach. Intell.* 27(8), 1226–1238 (2005)
12. Rakotomamonjy, A.: Variable selection using svm based criteria. *J. Mach. Learn. Res.* 3, 1357–1370 (2003)
13. Vapnik, V.N.: An overview of statistical learning theory. *IEEE Trans Neural Netw.* 10(5), 988–999 (1999)